The confidence model may be trained using a logistic regression model. In particular, the probability, p, that the recognition is acceptable is modeled as:

$$\log\left(\frac{p}{1-p}\right) = B^T X,$$

where B is the vector of model parameters to be trained, T denotes the transpose of the vector B, and X is the vector of predictor scores. In other implementations, generalized additive models or neural networks may be substituted for the logistic regression model in order to capture nonlinearities in the predictor terms.

Selection of the predictors (step **205**) and training of the confidence model (**210**) only need to be done once in training the confidence-based system. Typically, this training is done before the system is deployed although the system may be retuned or adapted after deployment. Thereafter, when an utterance is received (step **215**) and recognized (step **220**), the speech recognition system (or the confidence-based system) extracts values of the predictors (step **225**) and provides those values to the confidence model to generate the confidence estimate (step **230**).

The threshold against which the confidence estimate is compared may be tuned using additional training data. In particular, the threshold may be set to a value that optimizes the trade-off between passing a desired portion of acceptably recognized utterances while not passing too many unacceptably recognized utterances.

The described systems, methods, and techniques may be implemented in digital electronic circuitry, computer hardware, firmware, software, or in combinations of these elements. Apparatus embodying these techniques may include appropriate input and output devices, a computer processor, and a computer program product tangibly embodied in a machine-readable storage device for execution by a programmable processor. A process embodying these techniques may be performed by a programmable processor executing a program of instructions to perform desired functions by operating on input data and generating appropriate output. The techniques may be implemented in one or more computer programs that are executable on a programmable system including at least one programmable processor coupled to receive data and instructions from, and to transmit data and instructions to, a data storage system, at least one input device, and at least one output device. Each computer program may be implemented in a high-level procedural or object-oriented programming language, or in assembly or machine language if desired; and in any case, the language may be a compiled or interpreted language. Suitable processors include, by way of example, both general and special purpose microprocessors. Generally, a processor will receive instructions and data from a read-only memory and/or a random access memory. Storage devices suitable for tangibly embodying computer program instructions and data include all forms of non-volatile memory, including by way of example semiconductor memory devices, such as Erasable Programmable Read-Only Memory (EPROM), Electrically Erasable Programmable Read-Only Memory (EEPROM), and flash memory devices; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and Compact Disc Read-Only Memory (CD-ROM). Any of the foregoing may be supplemented by, or incorporated in, specially-designed ASICs (application-specific integrated circuits).

A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the claims. For example, advantageous results still could be achieved if steps of the disclosed techniques were performed in a different order and/or if components in the disclosed systems were combined in a different manner and/or replaced or supplemented by other components. Accordingly, other implementations are within the scope of the following claims.

For example, some implementations may employ a series of thresholds such that, for example, utterances with the highest confidence are passed directly to the message recipient, utterances with somewhat lower confidence values trigger a confirmation/verification dialogue with the sender, and messages falling into the lowest confidence range are sent directly for human transcription. Some implementations may employ a threshold that is determined during routing and/or training. For example, the threshold may be determined based on noise in the system or the routing application.

What is claimed is:

1. A computer-based method of routing an utterance to a system, the method comprising:

receiving an utterance including two or more words;

processing the utterance using large-vocabulary continuous speech recognition to generate a string of text corresponding to the utterance;

generating a confidence estimate of the string of text corresponding to the utterance, the generating including:

selecting one or more predictors relating to the large-vocabulary continuous speech recognition, and

training a confidence model using the one or more predictors;

comparing the confidence estimate to a predetermined threshold;

if the confidence estimate satisfies the predetermined threshold, forwarding the string of text to the system; and

if the confidence estimate does not satisfy the predetermined threshold, forwarding information relating to the utterance to a transcriptionist.

2. The method of claim **1** further comprising if the confidence estimate does not satisfy the predetermined threshold, having the transcriptionist determine an acceptable string of text and forwarding the acceptable string of text to the system.

3. The method of claim **1** in which the confidence estimate indicates a probability that the string of text is an acceptable representation of the utterance.

4. The method of claim **1** in which the information relating to the utterance comprises one or more of: the utterance, the string of text corresponding to the utterance, and the generated confidence estimate of the string of text.

5. The method of claim **1** in which generating the confidence estimate comprises:

extracting values of the one or more predictors based on the received utterance; and

providing the extracted values to the confidence model to generate the confidence estimate.

6. The method of claim **1** further comprising:

comparing the confidence estimate to a second predetermined threshold; and

if the confidence estimate does not satisfy the first predetermined threshold and does satisfy the second predetermined threshold level: